Apascaritei, P., Demel, S., Radl, J. (2021). The Difference Between Saying and Doing: Comparing Subjective and Objective Measures of Effort Among Fifth Graders. *American Behavioral Scientist*, pp. 1-23.

DOI: https://doi.org/10.1177/0002764221996772

THE DIFFERENCE BETWEEN SAYING AND DOING:

COMPARING SUBJECTIVE AND OBJECTIVE MEASURES OF EFFORT AMONG 5TH GRADERS

Paula Apascaritei[1], Simona Demel[1] & Jonas Radl[1,2]

Universidad Carlos III de Madrid

WZB Berlin Social Science Center

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 758600).

Abstract

The first goal of this study is to examine the capacity of prominent survey-based effort proxies to predict real effort provision in children. Do children who "talk the talk" of hard work also "walk the walk" and make costly effort investments? The second goal is to assess how objective and subjective effort measures are related under two conditions: intrinsic (non-incentivized) motivation and extrinsic (incentivized) motivation. We measure objective "real" effort using three tasks and subjective self-reported effort using four psychological characteristics (conscientiousness, need for cognition, locus of control and delay of gratification) to understand to what extent material incentives affect the cognitive effort of children with different self-reported personalities. Data stem from real-effort experiments carried out with 420 students attending the 5th grade of primary school in Madrid, Spain. We find that some of the subjective and objective effort measures are positively correlated, yet the power of personality to predict real effort is only moderate, and more so in the extrinsic than the intrinsic motivation condition. In particular, need for cognition and conscientiousness are the most relevant correlates of objective effort. Overall, we find there is a big difference between saying and doing when it comes to exerting effort, and this difference is even larger when there are no direct material incentives in place to reward effort provision.

Background and motivation

Personality traits have been argued to be crucial for socio-economic achievement (Kautz, Heckman, Diris, Ter Weel, & Borghans, 2014), making them an important factor in understanding social mobility (Doren & Grodsky, 2016; Hsin & Xie, 2017; also see Holtmann, Menze & Solga in this issue). There is evidence that personality traits are more predictive of relevant life outcomes than IQ, especially when it comes to health (Borghans, Golsteyn, Heckman, & Humphries, 2016). A recent meta-analysis suggests that personality is as decisive for achievement as family background (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Moreover, childhood personality is especially malleable, rendering social interventions especially effective at an early age (Kautz et al., 2014). However, the precise mechanisms by which childhood personality traits affect life outcomes remain nebulous as the proper way to measure personality traits continues to be disputed (Lechner, Danner, & Rammstedt, 2019; Rimfeld, Kovas, Dale, & Plomin, 2016). Similarly, there are concerns about social desirability bias in survey-based personality measures (cf. Radl & Miller in this issue) and recent evidence suggests that the long-term impact of psychological traits such as delay of gratification is less robust when fully accounting for unobserved heterogeneity (Watts, Duncan, & Quan, 2018).

We propose that theoretically, the impact of children's characteristics like conscientiousness or delay of gratification should be largely mediated by actual behavior, particularly effort (Cubel, Nuevo-Chiquero, Sanchez-Pages, & Vidal-Fernandez, 2016). This implies that there may be certain personality profiles that are more willing to engage in effort (Cassar & Meier, 2018). Children with such a profile would also have higher self-control and be more likely to keep their goals in mind, thus achieving superior educational outcomes. Hence, it is important to understand the ability of (subjectively measured) personality to predict (objectively measured) real effort. Do children who "talk the talk" of hard work also "walk the walk" and make costly effort investments?

Furthermore, previous studies have shown that individuals with certain personality traits may respond differently to incentives (Borghans, Meijers, & Ter Weel, 2008; Cubel et al., 2016). Based on standard economic theory, we propose that incentivizing task performance has a positive effect on exerted effort. Moreover, given individual heterogeneity, we further theorize that children with specific personality configurations will respond stronger to incentives.

This paper presents first evidence from a large-scale study on cognitive effort carried out in Spain. We examine the capacity of four prominent psychological scales (conscientiousness, need for cognition, locus of control, and delay of gratification) to explain real effort provision. Participants performed three different tasks: the Slider (Gill & Prowse, 2012), AX (Gonthier, Macnamara, Chow, Conway, & Braver, 2016) and Simon tasks (Cespón, Galdo-Álvarez, & Díaz, 2016) in two conditions. In the extrinsic motivation condition, they were offered material incentives and in the intrinsic motivation condition, performance was not linked to rewards.

Our contribution is threefold. First, we show to which extent the selected survey-based effort proxies predict observed effort intensities under laboratory conditions.[1] Second, we test to which extent effort provisions depend on whether material incentives are provided for task performance. Third, we assess whether children with certain personality profiles are more susceptible to incentive effects.

Literature review and conceptual considerations

[1] For stylistic purposes, we use different terms (personality scales, subjective effort measures, survey-based effort proxies) to refer to the selected personality traits that are conceived as relevant for effort. We call them subjective, because they are purely self-assessed whereas objective effort (also referred to as exerted or real effort) is measured through observable behaviors.

Real effort

In economics, effort is usually defined in opposition to ability. Mostly, it refers to how hard people try in a given task, in contrast to the skill needed for its completion. Real-effort tasks "measure the behavior of participants given specific observable tasks" (Charness, Gneezy, & Henderson, 2018, p. 75). While some effort tasks depend on cognitive capacity, such as performing mathematical operations, other tasks, such as sorting gravel by color into pots (Demel, Barr, Miller, & Ubeda, 2019) or adjusting the position on a digital slider (Gill & Prowse, 2012), are considered mostly independent of it.

In psychology, effort is a subjective phenomenon referring to "the degree of engagement with demanding tasks" employing executive functions (Westbrook & Braver, 2015, p. 396; Duckworth et al., 2019). Cognitive effort is similar to the concept of "cognitive load" (Paas, Tuovinen, Tabbers, Van Gerven, 2003), but also explicitly includes motivation. Executive functions are "processes that control and regulate thought and action" (Friedman et al., 2006, p. 172), and are underpinned by processing speed, attention, impulse control, and similar characteristics. They enable flexible thinking and inhibitory control and allow individuals to exert self-control during effortful tasks (Miyake & Friedman, 2012).

Combining the approaches to effort from psychology and economics, in this study, objective effort is considered as an individual's performance in real-effort tasks demanding minimal ability, whilst engaging various executive functions. After controlling for fluid intelligence, this approach provides an objective observable assessment of an individual's effort behavior in a laboratory setting.

Personality traits as effort proxies

Previous research has assessed the relationship between self-reported personality traits and effort, as measured by task performance on executive functions, but the amassed evidence is inconsistent. In some studies, personality and effort in executive function tasks are positively correlated (Duckworth & Kern, 2011; Malanchini, Engelhardt, Grotzinger, Harden, & Tucker-

Drob, 2019), while in others this relationship is mostly indistinguishable from zero (Buchanan, 2016; Unsworth et al., 2009). The meta-analysis by Duckworth and Kern (2011) reports correlations among executive function tasks and various self-reported personality traits that are weak, yet, statistically significant. Unsworth and colleagues (2009) find almost no correlation between self-reported conscientiousness and executive functions. Buchanan (2016) describes how certain personality dimensions are correlated to self-reported executive functions, but not particularly to actual performance on executive function tasks.

In this study, we consider four different personality traits[2]: a) conscientiousness, b) need for cognition, c) locus of control, and d) delay of gratification. These particular scales are chosen because the theoretical concepts they are designed to capture are directly relevant to effort. Although each scale emphasizes additional aspects that go beyond effort as such, close associations with real effort should be expected.

One of the most widely used models of personality is the Big Five model specifying five personality dimensions: extraversion, agreeableness, conscientiousness, neuroticism and openness (John, Naumann, & Soto, 2008). Among these traits, *conscientiousness* "describes socially prescribed impulse control that facilitates task- and goal-directed behavior" (John et al., 2008, p. 138). Thus, conscientiousness is the Big Five personality factor directly related to self-control and goal maintenance. Similar to the related concept of grit, it is seen as a determinant of the attainment of goals that require effortful self-control, such as academic achievement (Duckworth et al., 2019).

Need for cognition is a personality characteristic used in psychology to assess the "tendency to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116), referring to the intrinsic motivation of individuals towards cognitive effort. Individuals with high need for cognition are, therefore, more likely to engage in cognitive effort willingly (Inzlicht, Shenhav, & Olivola, 2018; Sandra & Otto, 2018). Indeed, need for cognition has been shown to predict academic performance in children and high school students (Keller et al., 2019). Malanchini and

[2] Please note that throughout the paper, the terms noncognitive/soft skills, personality/psychological traits, self-reported effort, effort proxies and subjective effort refer to these four personality variables.

colleagues (2019) pool various measures to create "superfactors" of openness, which is composed of multiple measures including need for cognition, and also of conscientiousness, which is composed of grit and related scales. These "superfactors" of openness and conscientiousness have a moderate association with executive function performance. A different study reports that need for cognition predicts cognitive effort, and concludes that it is an important component of intrinsic motivation (Sandra & Otto, 2018).

In addition, motivation is related to an individual's perception of control. This perception is captured by *locus of control* (Rotter, 1990): if internal attributions are made, one believes that events and experiences depend on one's choices and actions, which encourages one to actively pursue goals. If external attributions are made, one believes that outcomes depend on uncontrollable elements, such as luck or fate (Rotter, 1990). An internal locus of control is associated with positive outcomes that require effortful self-discipline, such as educational achievement and better health (Murasko, 2007) as well as more inequality acceptance (see Aguiar, Álvarez & Miller in this issue). A related construct, meritocratic beliefs, refers to people's casual conceptions about social mobility in society at large (Mijs, 2019).

Locus of control has been linked to various executive functions, from working memory to impulse control, based on the hypothesis of shared biological antecedents (Declerck, Boone, & De Brabander, 2006). Individuals with an internal locus of control have superior self-regulation abilities (Declerck et al., 2006). Internality of locus of control is not very different to high self-esteem and low neuroticism (Judge, Erez, Bono, & Thorensen, 2002), making it fundamental in achieving results. Even so, some researchers find that locus of control does not predict accuracy of responses or reaction time in executive function tasks (Muir et al., 2019).

Another aspect of personality related to self-regulation is *delay of gratification*. It allows individuals to be patient and to keep their focus on long-term goals instead of momentary pleasure. One of the early mechanisms proposed to capture this measure was the famous marshmallow test (Mischel, Ebbesen, & Raskoff-Zeiss, 1972). In our survey, children were asked to answer the question: "Imagine someone wants to give you a gift. Would you prefer receiving one gift today or two next week?" (Blossfeld & Roßbach, 2019). Although the reward is hypothetical in our case, studies show that real and hypothetical rewards in intertemporal choice decisions yield comparable results (Bickel, Pitcock, Yi, & Angtuaco, 2009; Brañas-Garza, Jorrat, Espín, & Sánchez, 2020). Delay of gratification correlates with positive outcomes in life, such as better health or less crime (Moffitt et al., 2011). Similar to the related concept of patience, delay of gratification increases with age, and correlates with educational outcomes (Sutter, Zoller, & Glätzle-Rützler, 2019). In economics, there is a closely related literature on time preferences (or future discounting) (see the contribution by Bortolotti et al. in this issue).

Failing to account for cognitive ability when studying the relation between personality and executive functions can confound results (Malanchini et al., 2019; Unsworth et al., 2009). Fluid intelligence has a separate effect on academic ability even after controlling for processing speed (Malanchini et al., 2019). Researchers also found that the relation between executive functions and intelligence in early childhood ranges from low to moderate depending on the executive function considered (Willoughby, Blair, Wirth, & Greenberg, 2010). Fluid intelligence is positively related to inhibition, shifting and working memory (Friedman et al., 2006, 2008). For this reason, fluid intelligence is an important factor to be considered when studying the relation between personality traits and effort.

Taken together, there is mixed evidence of the multi-faceted relationship between personality and effort. Therefore, there is a need for an in-depth assessment of whether personality scales related to self-regulation are indeed reliable indicators of effort as behavior. To this end, we select established measures, such as conscientiousness, locus of control and delay of gratification, which are often used as proxies of effort, and also a less studied measure - need for cognition – that is closely linked theoretically. Overall, we expect that these subjective measures of effort, are positively associated with objective effort, measured as task performance.

Hypothesis 1: The correlation between subjective and objective measures of effort is positive.

Rewarding effort

Cognitive effort is costly. An approach integrating economic and psychological insights posits that individuals will engage in cost-benefit analyses to choose optimal cognitive effort engagement (Westbrook & Braver, 2015). The costs of effort include biological (energy expenditure), emotional (boredom) and opportunity costs (of foregoing other rewarding activities) (Kurzban, 2016; Sandra & Otto, 2018). The benefits of effort include the emotional pleasure found in achieving goals and solving difficult tasks, especially among intellectually curious individuals (Ryan & Deci, 2000; Sandra & Otto, 2018; Segal, 2012).

One way to increase the benefit side of this cost-benefit analysis is through piece-rate rewards (Sandra & Otto, 2018). Based on standard economic theory, contingent rewards have a motivational effect on individuals (Benabou & Tirole, 2003). Once extrinsic rewards are tied to the successful realization of a task that has a clearly defined and measurable goal, individuals will strive to increase their output.

Extrinsic rewards boost performance by facilitating cognitive control during task completion (Van Steenbergen, Band, & Hommel, 2009). When incentives are attached to performance in executive function tasks, younger individuals respond faster and make fewer mistakes (Schmitt, Ferdinand, & Kray, 2015).

Hypothesis 2: Objective effort is higher in the extrinsic condition as compared to the intrinsic condition.

However, evidence also shows that extrinsic rewards may crowd out intrinsic motivation (Benabou & Tirole, 2003; Sandra & Otto, 2018). Indeed, individuals respond differently to incentives depending on their personality traits, for example, intellectually curious individuals are less responsive to incentives (Borghans et al., 2008). Similarly, conscientiousness has also been shown to positively affect productivity (Cubel et al., 2016).

Based on the literature briefly summarized above, we expect individuals with a perception of control to have superior performance in the extrinsic condition. If they believe that rewards depend on their own effort and if they have the ability to control their impulses, they will respond to the piece-rate incentives and engage effortful self-regulation for task performance. In contrast, individuals who have intrinsic motivation when it comes to orderliness, focus and consistency, and who find pleasure in thinking will also make an effort in the absence of incentives. To sum up, we propose that:

Hypothesis 3: Conscientiousness (H3a) and need for cognition (H3b) have stronger associations with intrinsic effort, as compared to (H3c) locus of control and (H3d) delay of gratification, which conversely have stronger associations with extrinsic effort.

Data and methods

Experimental design

Real-effort experiments were carried out from October 2019 to February 2020. The target population was 5th grade primary school students in the broader urban area of Madrid, Spain. Schools were randomly selected from a sampling frame stratified by neighborhood income quartile and school type (public, private, mixed). We approached school administrations by sending them a letter of invitation to participate in our project and followed up via telephone. The study was framed as an extracurricular activity where children visit the university campus and learn about higher education, while also performing fun activities and tasks. Each parent signed a consent form and data protection agreement prior to child participation in the experiment. The entire experimental process was reviewed by the ethics board and data protection officer at University Carlos III of Madrid, and approved. The response rate at the school level was approximately 30%. The sample comprises of 420 students, the vast majority between 10 and 11 years of age.[3]

All participants carried out three real-effort tasks chosen to address the differences between the interdisciplinary approaches to defining and understanding effort and taking into account that measuring effort involves "task impurity" (Miyake & Friedman, 2012). Thus, we proposed a multidimensional assessment of effort tasks, with three tasks focusing on different executive functions. By assessing these various types of executive functions, as well as controlling for cognitive ability, our framework should allow for the mitigation of variance not relating to actual effort.

The first of these tasks was the "Slider Task", one of the most popular real-effort tasks used in experimental economics, focusing on processing speed and goal maintenance (Gill & Prowse, 2012). In this task, participants saw 48 lines on their computer screens with sliders positioned at the left extreme of the line. The objective of the task was to move as many sliders as possible to the middle of the line in 2 minutes. The countdown timer was visible at the top of each computer screen.

The "AX-Continuous Performance Task" (AX-CPT) is used in psychology to assess switching between proactive and reactive cognitive control (Gonthier, Macnamara, Chow, Conway, & Braver, 2016). In this task, a trial consists of a pair of letters appearing one at a time. If participants saw the letter "A" followed by the letter "X" (or the sequence A – X), they were instructed to press the blue key on their keyboard, if they saw any other sequence, they were told to press the orange key.[4]

The "Simon Task" is a cognitive psychology task focused on inhibition and attention (Cespón, Galdo-Álvarez, & Díaz, 2016). In this task, participants saw an arrow on their computer screen, pointing either left or right. These arrows randomly appeared on different

[3] One participant is excluded from the analyses of the extrinsic condition due to technical issues experienced during the experiment.
[4] In order to make it easier for the participants, a blue sticker was placed on the "L" key and an orange sticker was placed on the "S" key and these keys were referred to by their color instead of their letter. Please see the Appendix for additional details regarding the tasks and the experimental design.

sides of the screen (left, middle, or right). If the arrow pointed left, participants were instructed to press the orange key on the left-hand side of the keyboard, and if the arrow pointed right, they were asked to press the blue key on the right-hand side of the keyboard, regardless of the arrow's position on the screen. Arrows were presented until a response was registered or timed out.

The objective of both the AX-CPT and the Simon task was to respond correctly to as many trials as possible in 120 seconds. For both tasks, subjects were told that the faster they responded, the more trials they would see, but were cautioned not to sacrifice accuracy for the sake of speed.

The experimental setup included an intrinsic and an extrinsic condition. In the intrinsic condition, participants did not receive any points for responding correctly to trials, yet were instructed to try their best. In the extrinsic condition, material incentives were provided in the form of points to be accrued for successful task completion. Children were informed they could take home the selected toys at the end of the day when the extrinsic condition was introduced. At the end of the experimental session, students could spend the accumulated points on a menu of toys (ranging from a widget spinner to a Lego set).

Many studies have found that including a leisure task in the experimental design uncovers incentive effects which are otherwise undetectable (Araujo et al., 2016; Corgnet, Hernán-González, & Schniter, 2015; Goerg, Kube, & Radbruch, 2019). Without a leisure task alternative, participants might exert effort in the lab simply because there is no opportunity cost of working on the tasks. This is not reflective of effortful tasks in the real world, since people almost always have the option to spend their time doing something else. Therefore, we decided to include leisure tasks in our experimental design. At the beginning of each 2-minute round, participants had to choose between completing the task or playing one of two mouse-based leisure tasks (a puzzle or a ball-bouncing game).

The order in which participants received the tasks varied by group. Participants always performed the first task under the intrinsic scheme first, and then carried out the first task, along with the remaining two tasks, under the extrinsic scheme, as can be seen in Table 1 . Thus, for each participant we have one observation of her objective effort measure under the intrinsic condition and three observations of her objective effort measure under the extrinsic condition. Data in Table 2 further illustrate the sample size for each task order. To sum up, 164 (75+89) subjects performed the slider task in the intrinsic condition, 186 (71+115) performed the AX task, and, finally, 70 subjects performed the Simon task for the intrinsic condition. Note that all subjects performed all of the three tasks in the extrinsic condition, thus rendering a sample size of 420 for this condition.

Table 1: Outline of the experimental setup

Outline	Time
Task 1:	
• Instructions + Control Questions + Practice Round(s)	Approx. 15 min
Leisure task:	
• Instructions + Practice Rounds	1 round of 1.5 min for each leisure task
Task 1:	
• Intrinsic condition	2 rounds of 2 min each
• Extrinsic condition	2 rounds of 2 min each
Task 2:	
• Instructions + Control Questions + Practice Round(s)	Approx. 10 min
• Extrinsic condition	2 rounds of 2 min each
Task 3:	
• Instructions + Control Questions + Practice Round(s)	Approx. 10 min
• Extrinsic condition	2 rounds of 2 min each

Table 2. Sample size by task order

Task order	Sample size
Slider-AX-Simon	75
Slider-Simon-AX	89
AX-Slider-Simon	115
AX-Simon-Slider	71
Simon-AX-Slider	70
Total	420

After the experimental session, children had a break that lasted approximately 45 minutes, where they visited the university campus, had a snack, and had time to play freely.

After this leisure break, they returned to the laboratory and performed a 5-minute fluid intelligence test followed by a survey with questions about their socio-demographic status and effort-related psychological scales and personality traits, among others.

Measures

Conscientiousness is a measure of three items from the Pictorial Personality Traits Questionnaire for Children (Mackiewicz & Cieciuch, 2016). *Need for cognition* is a measure of four items assessing enjoyment and engagement in knowledge pursuit (Beißert, Köhler, Rempel, Beierlein, 2014). *Locus of control (LOC)* is measured through 6 items (Jakoby & Jacob, 1999). Three items reflect internal locus of control, and three items reflect external locus of control. Both subscales are included as separate measures with higher values representing internal control.

Conscientiousness, need for cognition and locus of control were each measured on 5-point labeled Likert scales. Items were averaged to obtain the final measures and standardized prior to estimation. The dichotomous measure for *delay of gratification* is used in the German National Educational Panel Study (Blossfeld & Roßbach, 2019). All personality measures have been previously validated in the original studies mentioned above and have been found to be reliable. Items depart somewhat from the original scales due to translation into Spanish and simplification into a child-friendly form (see Table B1 in Appendix).

Cognitive ability is measured by fluid intelligence using Raven's progressive matrices test (Raven, Court & Raven, 1996). Children were asked to complete as many matrices as possible in 5 minutes. This measure was standardized for use in the regression models. Other variables included as controls in the regression analyses are the age of the participants measured in months and their sex. Since participants had to use a mouse in the Slider task, we determine children's familiarity with the use of keyboards and computer mice on a 4-point scale. Socio-economic status is controlled for via dummy variables measuring the neighborhoods income where the children's school is located (in quartiles within the overall income distribution of the Madrid region). Task order is controlled for via dummy variables.

We used multiple imputation via chained equations for missing values. In addition to the variables used in the analytical models, indicator variables containing teacher observations for each student regarding potential learning or behavioral difficulties were also included in the imputation procedure. While descriptive analyses are performed using non-imputed data only, 25 sets of imputed data are employed in the multivariate model estimations.

Model design

We estimate multilevel linear mixed models, with pupils nested in 18 school classes, clustering standard errors at the class level, to analyze whether survey-based effort proxies predict real effort provision. The model is:

$$y_{ij} = \alpha + \beta x_{ij} + u_j + \varepsilon_{ij}, \tag{1}$$

where y_{ij} is the effort measure, x_{ij} is a vector with the survey-based effort proxies and control variables of student i in class j, and β is a vector of the corresponding fixed effects. Furthermore, u_j is the random effect of class j, a residual component capturing the unobservable values characterizing the class, shared by all its students. Therefore, $\alpha + u_j$ is the random intercept of class j. Lastly, ε_{ij} are the residual components at the student level, assumed to be independent among students and among u_j.

The effort measure in the intrinsic incentive scheme is the number of total correct responses standardized within each task. Since each participant has three observations of effort in the extrinsic incentive scheme, we sum the standardized values of the number of total correct responses for each of the three tasks, and re-standardize the measure. In both incentive schemes, effort measures for participants who played the game instead of doing the task were coded as having zero total correct responses in that round. In the Online Appendix, we show model estimates for two additional outcomes related to the AX and Simon tasks: (i) the response accuracy and (ii) the average reaction times of the correct responses, as commonly used in the psychology literature.

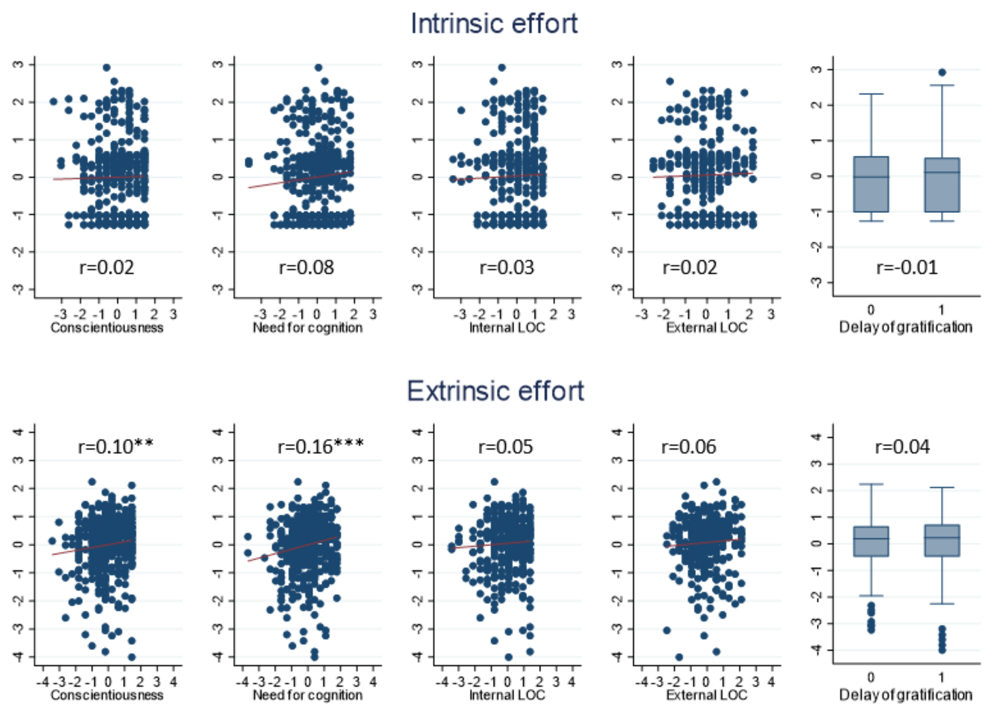
Results

In Figure 1, correlations between real effort in the intrinsic and the extrinsic motivation conditions and personality traits are presented. The associations between the objective and subjective measures of effort present a heterogeneous pattern: only need for cognition and conscientiousness are significantly associated with performance in the extrinsic scheme. These findings lend only very partial support to H1, especially considering the surprisingly low magnitudes of the coefficients.

Figure 2 shows the average number of total correct responses in one round in the intrinsic versus extrinsic incentive schemes within each task. The samples in the extrinsic schemes are restricted to the participants who did the same task as in the intrinsic scheme, therefore our total sample is divided across the three tasks.[5] The main difference between the

[5] In the first two bars, or in the "subjects doing leisure tasks included" version, the total correct responses are averaged across both rounds, with participants who played games coded as having zero total correct responses in that round. In the second two bars, or in the "subjects doing leisure tasks excluded" version, we removed the effect of game playing and only analyzed the effort in the tasks. This was done by only studying the total correct in one round of each incentive scheme. If participants played a game in one of the rounds, the number of total correct responses from the other round was included. If no game was played, the total correct was averaged across both rounds, and only if participants played games in both rounds were they excluded.

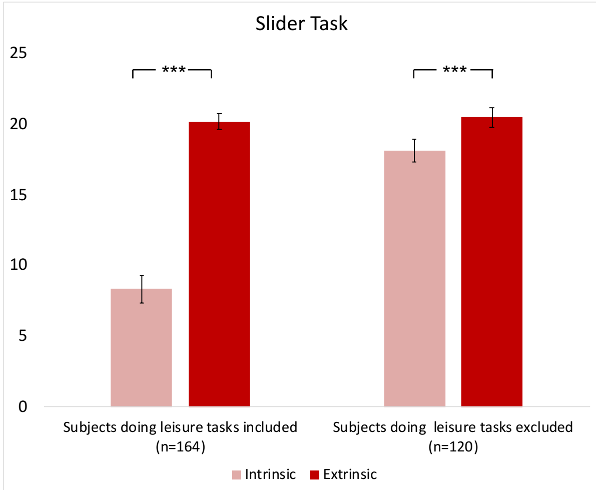
Figure 1. Correlations of main variables

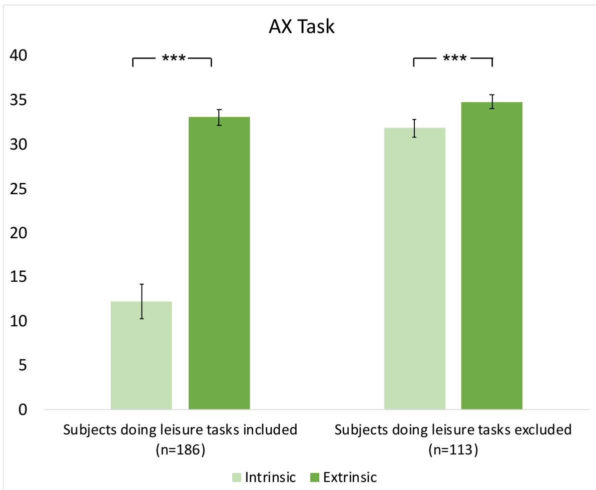


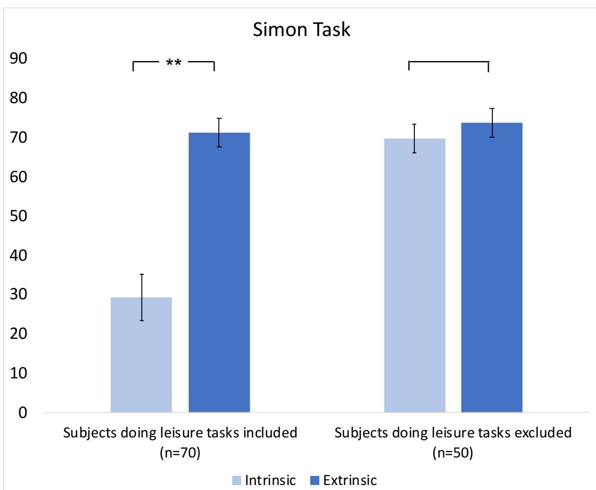
Note: Correlations reported in each panel. Correlations with delay of gratification are point-biserial correlations. Delay of gratification coded as follows: 0 = prefers one gift today, 1 = prefers two gifts in one week. * *p < 0.10,* ** *p < 0.05,* *** *p < 0.01*

intrinsic and extrinsic scheme is driven by choosing to play a game or not, with far more participants playing a game in the intrinsic scheme than in the extrinsic scheme. This difference is statistically significant at the 5% significance level or better, across the three tasks. When we compare only the level of effort in the tasks by looking at the version without subjects doing leisure tasks, we can see that while the difference in effort between the incentive schemes is much smaller, it is still significant at the 1% significance level for the Slider and AX tasks. However, the difference in effort between the incentive schemes is not statistically significantly different for the Simon task, which could be due to the small sample size. In sum, we find a clear incentivization effect on real effort, in line with H2.

Figure 2. Difference in effort by incentive scheme for each task







Note: For comparability, all scores refer to one round. The difference between the incentive schemes was calculated using a paired t-test clustering standard errors at the class level. The error bars refer to the standard error, after clustering standard errors at the class level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To check if this increase in total correct responses in the extrinsic scheme could be due to a learning effect, we analyze the change in either the total correct responses or the accuracy between rounds. As can be seen from Table D1 in the Appendix, we can safely say that participants make more of an effort in the extrinsic scheme than in the intrinsic scheme (apart from in the Simon task), and that this is not driven by a learning effect.

Table 3 presents regression results from the multi-level model in equation (1) for the intrinsic effort measure. Personality variables are introduced separately in models (1) to (4), and jointly in model (5). Examining the associations between objective and subjective effort, we only find a (marginally) statistically significant effect for need for cognition in model (2), confirming H3b.

Table 3. Results of analyses for the intrinsic condition

	(1) Intrinsic effort	(2) Intrinsic effort	(3) Intrinsic effort	(4) Intrinsic effort	(5) Intrinsic effort
Age in months	0.017^{*}	0.017^{*}	0.017^{*}	0.017^{*}	0.017^{*}
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Boy	-0.101	-0.098	-0.084	-0.091	-0.098
	(0.098)	(0.098)	(0.100)	(0.100)	(0.102)
Fluid intelligence	0.039	0.024	0.032	0.036	0.026
	(0.032)	(0.032)	(0.032)	(0.032)	(0.035)
Mouse use	0.056^{*}	0.056^{*}	0.060^{**}	0.057^{*}	0.061^{*}
	(0.031)	(0.031)	(0.030)	(0.032)	(0.034)
Conscientiousness	0.053				0.039
	(0.061)				(0.063)
Need for cognition		0.069^{*}			0.060
		(0.040)			(0.037)
Internal LOC			0.018		-0.005
			(0.046)		(0.047)
External LOC			0.031		0.023
			(0.053)		(0.056)
Delay of gratification				-0.029	-0.032
				(0.107)	(0.109)
Constant	-2.106^{**}	-2.139^{**}	-2.125^{**}	-2.081^{**}	-2.085^{**}
	(0.984)	(1.003)	(0.981)	(0.964)	(0.994)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0.314^{***}	0.316^{***}	0.315^{***}	0.315^{***}	0.315^{***}
	(0.052)	(0.052)	(0.053)	(0.052)	(0.052)
sd(ε_{ij})	0.925^{**}	0.924^{***}	0.925^{***}	0.926^{***}	0.922^{***}
	(0.028)	(0.028)	(0.027)	(0.027)	(0.028)
Observations	420	420	420	420	420

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 is the analog of Table 3 for the extrinsic condition, where participants were rewarded points for each trial correctly answered, and each slider correctly placed. In terms of subjective effort, both conscientiousness (although only marginally) and need for cognition are associated with objective effort in models (1) and (2), respectively. Once introduced jointly in model (5), only need for cognition preserves a marginally statistically significant, yet diminished, association. These results are contrary to our expectations in H3c and H3d, as the effect of locus of control and delay of gratification do not reach statistical significance neither in the intrinsic nor in the extrinsic condition. Notably, fluid intelligence is strongly associated with task performance in the extrinsic condition. The fact that this was not the case in the intrinsic condition suggests that rewards effectively incentivize effort among more intelligent children.

Table 4. Results of analyses for the extrinsic condition

	(1) Extrinsic effort	(2) Extrinsic effort	(3) Extrinsic effort	(4) Extrinsic effort	(5) Extrinsic effort
Age in months	0.009	0.009	0.008	0.009	0.010
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Boy	0.332^{***}	0.343^{***}	0.363^{***}	0.350^{***}	0.326^{***}
	(0.093)	(0.094)	(0.093)	(0.097)	(0.096)
Fluid intelligence	0.251^{***}	0.221^{***}	0.238^{***}	0.239^{***}	0.229^{***}
	(0.044)	(0.045)	(0.041)	(0.041)	(0.045)
Mouse use	0.081^{**}	0.083^{***}	0.088^{**}	0.076^{**}	0.077^{**}
	(0.033)	(0.032)	(0.037)	(0.035)	(0.036)
Conscientiousness	0.143^{*}				0.126
	(0.085)				(0.094)
Need for cognition		0.119^{**}			0.094^{*}
		(0.051)			(0.055)
Internal LOC			0.033		-0.005
			(0.055)		(0.056)
External LOC			0.036		0.013
			(0.064)		(0.066)
Delay of gratification				0.107	0.113
				(0.082)	(0.083)
Constant	-1.749	-1.807^{*}	-1.742	-1.923^{*}	-1.972^{*}
	(1.096)	(1.040)	(1.067)	(1.025)	(1.023)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.887^{**}	0.890^{**}	0.896^{**}	0.897^{**}	0.880^{**}
	(0.048)	(0.050)	(0.050)	(0.049)	(0.048)
Observations	419	419	419	419	419

Standard errors in parentheses
* *p < 0.10,* ** *p < 0.05,* *** *p < 0.01*

In the first half of Table 5, similar regressions are presented for both the intrinsic effort measure, as well as performance by task in the intrinsic condition.[6] The only statistically significant association, although only marginally significant, is that between the intrinsic effort measure and need for cognition. The lower half of Table 5 presents the estimation results for the extrinsic incentive scheme. The effect of conscientiousness and need for cognition on task performance are driven by the AX-CPT and Simon tasks. Additionally, delay of gratification has a marginally statistically significant effect for the Slider task performance, thus lending partial support to H3d. From Tables 4 and 5 it becomes clear that subjective effort is a better predictor of objective effort in the extrinsic condition. However, results should be interpreted cautiously when referring to specific task performance, as the sample size varies for each model in the intrinsic condition (see Table 2).

Table 5. Results of analyses by task for the intrinsic and extrinsic conditions

	(1) Effort	(2) Slider	(3) AX-CPT	(4) Simon
Intrinsic incentive scheme				
Conscientiousness	0.053	0.061	0.096	-0.072
	(0.061)	(0.108)	(0.095)	(0.140)
Need for cognition	0.069*	0.087	0.044	0.141
	(0.040)	(0.075)	(0.067)	(0.098)
Internal LOC	0.018	0.059	-0.023	0.077
	(0.046)	(0.082)	(0.056)	(0.158)
External LOC	0.031	0.094	-0.003	-0.015
	(0.053)	(0.067)	(0.075)	(0.205)
Delay of gratification	-0.029	0.029	-0.087	0.061
	(0.107)	(0.095)	(0.231)	(0.128)
Extrinsic incentive scheme				
Conscientiousness	0.143*	0.007	0.255**	0.050
	(0.085)	(0.090)	(0.107)	(0.082)
Need for cognition	0.119**	0.043	0.087**	0.133***
	(0.051)	(0.057)	(0.034)	(0.050)
Internal LOC	0.033	0.002	-0.001	0.076
	(0.055)	(0.053)	(0.055)	(0.048)
External LOC	0.036	0.074	0.033	-0.044
	(0.064)	(0.058)	(0.055)	(0.063)
Delay of gratification	0.107	0.115*	0.054	0.079
	(0.082)	(0.067)	(0.103)	(0.087)

Note: Models estimated with each scale separately (except internal and external locus of control together),

[6] For comparability purposes, results in column (1) of Table 5 show a stacked version of the results in models (1) through (4) in Tables 3 and 4. Columns (2), (3) and (4) present the equivalent coefficients for the task performance in the Slider, AX-CPT, and Simon tasks, respectively.

controlling for: sex, age, fluid intelligence, neighborhood income, task order, and mouse use (for models (1) and (2) only).
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We performed various sensitivity analyses (see Appendix) and found that the reported results remain qualitatively similar and are robust to treatment of missingness and outlier exclusion.[7] As an alternative outcome variable, we ran models on the difference in effort levels between the intrinsic and extrinsic incentive schemes. We found that none of the measures were statistically significantly different from zero, suggesting that the additional effort provided in the piece-rate condition was not systematically related to personality. Additionally, we considered interactive effects between sex and personality traits and find that, for most personality characteristics, there were no significant gender differences in the way they affect effort provisions, although subjective and objective effort seems to be somewhat more closely related among boys than among girls.

Discussion

What difference is there between self-reported and exerted effort? We report novel evidence from a large-scale study about the factors contributing to real effort provision among primary school children. We draw on novel data from laboratory experiments carried out in Spain with a sample of 420 subjects aged around 10-11 years. We assess the interplay between personality characteristics and incentivization in determining the effort displayed in three different cognitive tasks. The study delivers five contributions to the existing literature.

First, personality characteristics overall are demonstrated to have a limited capacity to explain children's real effort provision in the laboratory. Most bivariate correlations are not statistically significant and the largest measured correlation is only r=.17. After controlling for potential confounders, need for cognition and conscientiousness have robust significant effects

[7] We also checked if the number of times a child chooses to perform the task instead of play one of the games was related to one of the subjective measures. We found that need for cognition was the only measure that was statistically significant at the 5% level or better.

on real effort, while locus of control (internal and external) and delay of gratification do not. In terms of magnitude, the largest effect is found for conscientiousness: for example, an increase of one standard deviation in this personality trait increases average extrinsic effort by 14% of a standard deviation. These findings are broadly in line with previous meta-analyses indicating that executive functions and self-reported personality traits are moderately correlated, at best (Duckworth & Kern, 2011). This may indicate that many children are poor evaluators of their own effort engagement, or that their survey answers are subject to social desirability bias.

Second, providing material incentives – in the form of accrued points convertible into chosen toys – led to a significant increase in effort as compared to when no rewards were offered. Our study adds to the literature showing substantive incentive effects (Borghans et al., 2008; Dellavigna & Pope, 2018; Sandra & Otto, 2018; Schmitt et al., 2015), which underpin the proposed interpretation of task performance as effort.

Third, looking for the specific personality profiles related to the willingness to exert effort, the subjective personality scale showing the most consistent association with objective effort is need for cognition, in both the intrinsic and extrinsic condition. Sociologists and economists should take note of the proxy ability of need for cognition (Cacioppo & Petty, 1982), a personality scale that is relatively unknown outside of psychology. Unexpectedly, conscientiousness has a significant influence on effort provision only when rewards were in place, but not in the intrinsic condition. Neither delay of gratification nor internal or external locus of control can consistently predict students' effort.

Fourth, we propose a synthetic effort measure that averages performance across three standard tasks requiring the use of different executive functions – and control for the score on a matrix-based (non-incentivized) fluid intelligence test, in order to neutralize the residual influence of any cognitive abilities. When comparing the three tasks employed in the study, there were a few notable differences. The AX-CPT was slightly more affected by personality – specifically by need for cognition and conscientiousness – than the other two and, as additional analyses show, also less affected by fluid intelligence. All considered, we conclude that conscientiousness and need for cognition are more relevant for those tasks underpinned by effort invested in proactive and inhibitory control (AX-CPT task), while delay of gratification could be a better predictor of effort put into processing speed (Slider task).

Fifth, the study contains some specific lessons for cross-disciplinary dialogue. Real-effort tasks are employed in behavioral economics with the purpose of minimizing the influence of ability. However, performance in the Slider task, developed and mostly used by economists, is significantly correlated with fluid intelligence, but only when material incentives are in place (as is typically the case in experimental economics, but rarely in psychological studies). On the other hand, psychological personality scales such as conscientiousness and locus of control have stronger associations with incentivized than with non-incentivized effort, again at odds with the dominant disciplinary paradigm.

This study has several limitations. First, the experimental design limited the sample size for the intrinsic condition, requiring a careful interpretation of these results. This decision was driven by the need to avoid participant fatigue, especially since executive function performance can be affected by excessively long sessions. Thus, our sample is not balanced across the incentive schemes (we have more observations in the extrinsic condition), and it is also not balanced across the three tasks (we have fewer observations for the Simon task). Second, despite our effort to make the questionnaire as child-friendly as possible, we cannot rule out that null results were driven by measurement error. Third, since this is a cross-sectional study, we cannot assess the longitudinal effects of personality or the stability of effort across the life course, which should be assessed in future studies.

Conclusion

The main goals of this study were (a) to understand to which extent psychological scales, which are often used as proxies of effort, are able to predict children's observed effort intensities and (b) to test just how much the association between subjective and objective effort depends on material incentives. The key conclusion is that there is a big difference between saying and doing when it comes to exerting effort, and this difference is even larger when there are no direct material incentives in place to reward effort provision.

References

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., … Wilson, A. J. (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association*, *2*(1), 1–12. https://doi.org/10.1007/s40881-016-0025-7

Beißert, H., Köhler, M., Rempel, M., & Beierlein, C. (2014). Eine deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition [A German short-scale for measuring the construct need for cognition]. In *GESIS-Working Papers* (No. 32).

Benabou, R., & Tirole, J. (2003). Intrinsic and Extrinsic Motivation. *Review of Economic Studies*, *70*(3), 489–520. https://doi.org/10.1111/1467-937X.00253

Bickel, W. K., Pitcock, J. A., Yi, R., & Angtuaco, E. J. C. (2009). Congruence of BOLD response across intertemporal choice conditions: Fictive and real money gains and losses. *Journal of Neuroscience*, *29*(27), 8839–8846. https://doi.org/10.1523/JNEUROSCI.5319-08.2009

Blossfeld, H.-P., & Roßbach, H.-G. (Eds.). (2019). Education as a Lifelong Process. The German National Educational Panel Study (NEPS). *Edition ZfE – Zeitschrift Für Erziehungswissenschaft*, Vol. 3. Wiesbaden: Springer VS.

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354–13359. https://doi.org/10.1073/pnas.1601135113

Borghans, L., Meijers, H., & Ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic Inquiry*, *46*(1), 2–12. https://doi.org/10.1111/j.1465-7295.2007.00073.x

Brañas-Garza, P., Jorrat, D., Espín, A. M., & Sánchez, A. (2020). Paid and hypothetical time preferences are the same. *Working Paper*.

Buchanan, T. (2016). Self-report measures of executive function problems correlate with personality, not performance-based executive function measures, in nonclinical samples. *Psychological Assessment*, *28*(4), 372–385. https://doi.org/10.1037/pas0000192

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cassar, L., & Meier, S. (2018). Nonmonetary incentives and the implications of work as a source of meaning. *Journal of Economic Perspectives*, *32*(3), 215–238. https://doi.org/10.1257/jep.32.3.215

Cespón, J., Galdo-Álvarez, S., & Díaz, F. (2016). Cognitive control activity is modulated by the magnitude of interference and pre-activation of monitoring mechanisms. *Scientific Reports*, *6*(1), 1–11. https://doi.org/10.1038/srep39595

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74–87. https://doi.org/10.1016/j.jebo.2018.02.024

Corgnet, B., Hernán-González, R., & Schniter, E. (2015). Why real leisure really matters: incentive effects on real effort in the laboratory. *Experimental Economics*, *18*(2), 284–301. https://doi.org/10.1007/s10683-014-9401-4

Cubel, M., Nuevo-Chiquero, A., Sanchez-Pages, S., & Vidal-Fernandez, M. (2016). Do Personality Traits Affect Productivity? Evidence from the Laboratory. *Economic Journal*, *126*(592), 654–681. https://doi.org/10.1111/ecoj.12373

Declerck, C. H., Boone, C., & De Brabander, B. (2006). On feeling in control: A biological theory for individual differences in control perception. *Brain and Cognition*, *62*(2), 143–176. https://doi.org/10.1016/j.bandc.2006.04.004

Dellavigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*, *85*(2), 1029–1069. https://doi.org/10.1093/restud/rdx033

Demel, S., Barr, A., Miller, L., & Ubeda, P. (2019). Commitment to political ideology is a luxury only students can afford: A distributive justice experiment. *Journal of Experimental Political Science*, *6*(1), 33–42. https://doi.org/10.1017/XPS.2018.14

Doren, C., & Grodsky, E. (2016). What Skills Can Buy: Transmission of Advantage through

Cognitive and Noncognitive Skills. *Sociology of Education*, *89*(4), 321–342. https://doi.org/10.1177/0038040716667994
Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*(3), 259–268. https://doi.org/10.1016/j.jrp.2011.02.004
Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-Control and Academic Achievement. *Annual Review of Psychology*, *70*(1), 373–399. https://doi.org/10.1146/annurev-psych-010418-103230
Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x
Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201
Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469–503. https://doi.org/10.1257/aer.102.1.469
Goerg, S. J., Kube, S., & Radbruch, J. (2019). The effectiveness of incentive schemes in the presence of implicit effort costs. *Management Science*, *65*(9), 4063–4078. https://doi.org/10.1287/mnsc.2018.3160
Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R. A., & Braver, T. S. (2016). Inducing proactive control shifts in the AX-CPT. *Frontiers in Psychology*, *7*(1882). https://doi.org/10.3389/fpsyg.2016.01822
Hsin, A., & Xie, Y. (2017). Life-course changes in the mediation of cognitive and non-cognitive skills for parental effects on children's academic achievement. *Social Science Research*, *63*, 150–165. https://doi.org/10.1016/j.ssresearch.2016.09.012
Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007
Jakoby, N., & Jacob, R. (1999). Messung von internen und externen Kontrollüberzeugungen in allgemeinen Bevölkerungsumfragen. *ZUMA Nachrichten*, *23*(45), 61–71.
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five Trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). https://doi.org/10.1016/S0191-8869(97)81000-8
Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, *83*(3), 693–710. https://doi.org/10.1037/0022-3514.83.3.693
Kautz, T., Heckman, J. J., Diris, R., Weel, B. T., & Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. In *OECD Education Working Papers* (No. 110). https://doi.org/10.1787/19939019
Keller, U., Strobel, A., Wollschläger, R., Greiff, S., Martin, R., Vainikainen, M.-P., & Preckel, F. (2019). A Need for Cognition Scale for Children and Adolescents. *European Journal of Psychological Assessment*, *35*(1), 137–149. https://doi.org/10.1027/1015-5759/a000370
Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology*, *7*, 67–70. https://doi.org/10.1016/j.copsyc.2015.08.003
Lechner, C. M., Danner, D., & Rammstedt, B. (2019). Grit (effortful persistence) can be measured with a short scale, shows little variation across socio-demographic subgroups, and is associated with career success and career engagement. *PLoS ONE*, *14*(11), 1–29. https://doi.org/10.1371/journal.pone.0224814
Mackiewicz, M., & Cieciuch, J. (2016). Pictorial Personality Traits Questionnaire for Children (PPTQ-C)-A new measure of children's personality traits. *Frontiers in Psychology*,

7(498). https://doi.org/10.3389/fpsyg.2016.00498
Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2019). "Same but different": Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and Social Psychology*, *117*(6), 1164–1188. https://doi.org/10.1037/pspp0000224
Mijs, J. J. B. (2019). The paradox of inequality: income inequality and belief in meritocracy go hand in hand. *Socio-Economic Review*, 1–29. https://doi.org/10.1093/ser/mwy051
Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, *21*(2), 204–218. https://doi.org/10.1037/h0032198
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458
Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H. L., … Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108
Muir, A. M., Carbine, K. A., Goodwin, J., Hedges-Muncy, A., Endrass, T., & Larson, M. J. (2019). Differentiating electrophysiological indices of internal and external performance monitoring: Relationship with perfectionism and locus of control. *PLoS ONE*, *14*(10), 1–21. https://doi.org/10.1371/journal.pone.0219883
Murasko, J. E. (2007). A lifecourse study on education and health: The relationship between childhood psychosocial resources and outcomes in adolescence and young adulthood. *Social Science Research*, *36*(4), 1348–1370. https://doi.org/10.1016/j.ssresearch.2007.01.001
Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8
Raven, J. C., Court, J. H., & Raven, J. (1996). *Raven Matrices Progresivas*. Madrid: TEA Ediciones, S. A.
Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, *111*(5), 780–789. https://doi.org/10.1037/pspp0000089
Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science*, *2*(4), 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x
Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist*, *45*(4), 489–493. https://doi.org/10.1037/0003-066X.45.4.489
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020
Sandra, D. A., & Otto, A. R. (2018). Cognitive capacity limitations and Need for Cognition differentially predict reward-induced cognitive effort expenditure. *Cognition*, *172*, 101–106. https://doi.org/10.1016/j.cognition.2017.12.004
Schmitt, H., Ferdinand, N. K., & Kray, J. (2015). The influence of monetary incentives on context processing in younger and older adults: an event-related potential study. *Cognitive, Affective and Behavioral Neuroscience*, *15*(2), 416–434. https://doi.org/10.3758/s13415-015-0335-x
Segal, C. (2012). Working When No One Is Watching: Motivation, Test Scores, and Economic Success. *Management Science*, *58*(8), 1438–1457. https://doi.org/10.1287/mnsc.1110.1509
Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents – A first survey of experimental economics results. *European Economic Review*, *111*, 98–121. https://doi.org/10.1016/j.euroecorev.2018.09.004

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Thadeus Meeks, J., Keith Campbell, W., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences*, *30*(4), 194–200. https://doi.org/10.1027/1614-0001.30.4.194

Van Steenbergen, H., Band, G. P. H., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science*, *20*(12), 1473–1477. https://doi.org/10.1111/j.1467-9280.2009.02470.x

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes. *Psychological Science*, *29*(7), 1159–1177. https://doi.org/10.1177/0956797618761661

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, *22*(2), 306–317. https://doi.org/10.1037/a0018708

THE DIFFERENCE BETWEEN SAYING AND DOING:

COMPARING SUBJECTIVE AND OBJECTIVE MEASURES OF EFFORT AMONG 5TH GRADERS

Paula Apascaritei[1], Simona Demel[1] & Jonas Radl[1,2]

Universidad Carlos III de Madrid

WZB Berlin Social Science Center

APPENDIX

Table of contents

APPENDIX A: A Note on the Experimental Design

Data collection was organized in the form of daily field trips of entire school classes to the university campus, where a range of educational activities were organized around the experimental and survey sessions. In the experimental sessions, all participants carry out three real-effort tasks: the Slider task, the AX-CPT task and the Simon task. In the AX-CPT task, the start of each trial is marked by the presentation of a fixation point for 500 ms, followed by the first letter for 400 ms, a blank screen of 1200 ms, and finally, the second letter remains on the screen until the participant response is registered or times out. In the Simon task, arrows are presented until the participant response is registered or times out. For both the AX-CPT and the Simon task, the time-out for offering a response is set at 5000 ms. All elements are presented with a black font on a grey background. Participants are presented with two 2-minute blocks of trials (rounds) for each task. Between each two-minute round, participants could take a short break without leaving their workspace.

Instructions were given to the whole class, imitating a normal classroom setting. After the introduction of each task and new incentive scheme, participants had to answer control questions on their computer screens to ensure they correctly understood all relevant details. They also had the chance to try each task out in a practice round before starting the actual experimental rounds, receiving feedback about their performance after each trial for the Simon and AX tasks, which had a 1-minute practice round each. Since the Slider task involved the use of a computer mouse, the participants had 2 practice rounds of 2 minutes each, in line with the setup of Gill and Prowse (2012). After the practice round(s) of the first task, participants played each game for 90 seconds in order to familiarize themselves with the leisure tasks.

Participants who answered a control question incorrectly, or whose accuracy during the practice round was below 80% were flagged on our server. One of the experimenters then approached them and clarified any misunderstandings of either the task or incentive scheme with the participant.

Moreover, in the experimental set-up, the children carried out the experiment first, which ended with a tournament scheme (excluded from this paper's analysis), and the cognitive ability and survey component second. In the tournament scheme, they were told that the top three participants would have the honor of receiving a certificate. We are aware of Chen et al.'s (2020) findings that performing in an honor incentive had an effect on the subjects' self-reported measures of noncognitive skills. However, we are not concerned that the self-reported measures in our study suffered the same fate. Chen et al. (2020) administered the Big Five directly after their honor incentive, whereas, the children in our study were first taken outside for a campus tour and a snack, lasting approximately 45 minutes, before completing the survey component. Furthermore, the self-reported measures of interest were among the last questions to be asked in the survey.

Chen, Y., Feng, S., Heckman, J. J., & Kautz, T. (2020). Sensitivity of self-reported noncognitive skills to survey administration conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(2), 931–935.

Gill, D., & Prowse, V. (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, *102*(1), 469–503. https://doi.org/10.1257/aer.102.1.469

APPENDIX B: Personality Measures

Table B1. Items for personality measures

Variables	Items
Conscientiousness	1. When I get money from someone... I save it/I spend it right away.* 2. My room is… messy/orderly. 3. I do my housework... willingly/unwillingly.*
Need for cognition	1. I like exercises that make me think a lot. 2. I like challenges that I need to think about. 3. I prefer to think as little as possible.* 4. I just need to know the answer, I don't need to know the reason.*
Internal locus of Control External locus of control	1. I like taking responsibility and doing things myself. 2. I find it best to make decisions myself, rather than to rely on fate. 3. I usually find a way to overcome problems. 4. Success often depends more on luck than on effort.* 5. I often have the feeling that I have little control over what happens to me.* 6. When I make important decisions, I often look at what others have done.*
Delay of gratification	Imagine someone wants to give you a gift. Would you prefer receiving one gift today or two next week? • I would prefer receiving one gift today. • I would prefer receiving two gifts next week.

Notes: Items marked with an asterisk are reverse coded.

APPENDIX C: Sample Descriptive Statistics and Correlation Table

Table C1: Summary statistics of total correct answers by task and condition

Variables	Mean	Std. Dev.	Min.	Max.	Sample size
AX intrinsic	24.58	23.98	0	79	186
Slider intrinsic	16.64	13.10	0	55	164
Simon intrinsic	58.61	48.97	0	184	70
AX extrinsic	69.44	13.65	0	84	420
Slider extrinsic	38.84	12.11	0	79	420
Simon extrinsic	148.77	38.55	0	196	420

Table C2: Sample summary statistics

Variables	Mean	Std. Dev.	Min.	Max.	Sample size
Age in months	126.56	6.06	99	163	418
Boy	0.46	0.50	0	1	420
Conscientiousness	3.82	0.82	1	5	416
Need for cognition	3.70	0.73	1	5	417
Internal LOC	3.95	0.76	1.33	5	354
External LOC	3.16	0.88	1	5	313
Delay of gratification	0.66	0.47	0	1	418
Fluid intelligence	23.89	4.23	11	35	419
Mouse use	1.00	1.05	0	3	418

Table C3: Neighborhood income distribution

Variables	Mean	Std. Dev.	Min.	Max.	Sample size
First quartile	0.39	0.49	0	1	420
Second quartile	0.23	0.42	0	1	420
Third quartile	0.33	0.47	0	1	420
Fourth quartile	0.04	0.20	0	1	420

Table C4: Correlation table

Variables	(1) Intr. effort	(2) Slider intr.	(3) AX intr.	(4) Simon intr.	(5) Extr. effort	(6) Slider extr.	(7) AX extr.	(8) Simon extr.	(9) Consc.	(10) Need cogn.	(11) Int. LOC	(12) Ext. LOC	(13) Delay grat.	(14) Fluid intellig.	(15) Age in months	(16) Boy
Objective effort measures																
(1) Intrinsic effort	.															
(2) Slider intrinsic	.	.														
(3) AX intrinsic	.	.	.													
(4) Simon intrinsic	.	.	.	.												
(5) Extrinsic effort	0.17***	0.20***	0.17**	0.16	.											
(6) Slider extrinsic	0.11**	0.25***	0.03	-0.01	0.73***	.										
(7) AX extrinsic	0.20***	0.22***	0.25***	0.06	0.73***	0.29***	.									
(8) Simon extrinsic	0.08	-0.03	0.08	0.28**	0.75***	0.34***	0.34***	.								
Subjective effort measures																
(9) Conscientiousness	0.02	0.04	0.02	-0.05	0.10**	0.03	0.07	0.13***	.							
(10) Need for cognition	0.08	0.07	0.08	0.08	0.16***	0.07	0.11**	0.17***	0.27***	.						
(11) Internal LOC	0.03	0.04	0.00	0.06	0.05	0.02	0.01	0.09*	0.18***	0.23***	.					
(12) External LOC	0.02	0.09	0.03	-0.13	0.06	0.08	0.09*	-0.04	0.05	0.12**	0.08	.				
(13) Delay of gratification	-0.01	-0.00	-0.03	0.04	0.04	0.06	0.00	0.03	-0.00	0.01	-0.15***	0.05	.			
Other measures																
(14) Fluid intelligence	0.05	0.08	0.10	-0.13	0.28***	0.22***	0.20***	0.19***	-0.02	0.16***	0.04	0.16***	0.02	.		
(15) Age in months	0.10**	0.12	0.09	0.11	0.07	0.10*	0.02	0.03	-0.10**	-0.04	0.01	-0.03	-0.06	-0.03	.	
(16) Boy	-0.04	-0.14*	0.03	0.00	0.17***	0.19***	0.08	0.10**	0.04	0.04	-0.02	-0.12**	0.04	-0.06	0.05	.
(17) Mouse use	0.04	0.05	0.01	0.12	0.07	0.10*	0.00	0.06	0.05	-0.03	-0.01	-0.16***	0.11*	-0.09*	0.06	0.04

Note: Objective effort measures in the intrinsic condition have no reported correlations because each participant completed only one task in the intrinsic condition (see explanations in text and in Tables 1 and 2). Correlations with delay of gratification and sex are point-biserial correlations. Delay of gratification coded as follows: 0 = prefers one gift today, 1 = prefers two gifts in one week. Sex coded as follows: 0 = female, 1 = male.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX D: Learning Effects

To check if the increase in total correct responses in the extrinsic scheme with respect to the intrinsic scheme (as seen in Figure 2 in the article) could be due to a learning effect, we analyze the change in either the total correct responses or the accuracy between rounds. As can be seen from Table D1, there was a learning effect in the Slider task when comparing the two practice rounds and also between the second practice round and the first intrinsic round. On average, participants placed one more slider correctly in the following round. However, there is no further increase when comparing the two intrinsic scheme rounds or the two extrinsic scheme rounds. Furthermore, no differences are found between the rounds of the AX task. Lastly, while there appears to be a difference between the rounds of the extrinsic scheme for the Simon task, it goes in the opposite direction. Therefore, we can safely say that no learning effect is found.

Table D1: Learning effects by round

	Mean difference	p-value	Sample size
Slider task:			
Practice r1 vs practice r2	1.18	0.000	420
Practice r2 vs intrinsic r1	1.02	0.036	120
Intrinsic r1 vs intrinsic r2	0.09	0.915	32
Extrinsic r1 vs extrinsic r2	0.20	0.158	400
AX task:			
Practice round vs intrinsic r1 (accuracy)	2.41	0.383	114
Intrinsic r1 vs intrinsic r2	-0.03	0.982	29
Intrinsic r1 vs intrinsic r2 (accuracy)	0.64	0.782	29
Extrinsic r1 vs extrinsic r2	-0.15	0.507	399
Extrinsic r1 vs extrinsic r2 (accuracy)	-0.23	0.631	399

Simon task:			
Practice round vs intrinsic r1 (accuracy)	-1.30	0.639	50
Intrinsic r1 vs intrinsic r2	2.50	0.177	8
Intrinsic r1 vs intrinsic r2 (accuracy)	0.68	0.463	8
Extrinsic r1 vs extrinsic r2	-0.94	0.016	398
Extrinsic r1 vs extrinsic r2 (accuracy)	-1.27	0.006	398

Note: Paired t-tests were conducted for the number of total correct responses, or for the accuracy if specified in the brackets, clustering standard errors at the class level. The mean difference is calculated by subtracting the first term from the second, thus, a positive mean difference represents a learning effect. When a practice round was compared to intrinsic round 1, the latter refers to the first time they did the task in the intrinsic scheme, i.e., round 2 if they played a game in round 1. For the r1 vs r2 comparisons within the incentive schemes, only participants who did not play games in both rounds were considered, which accounts for the small sample sizes in some cases. For the AX and Simon tasks, since the practice round was only 1 minute and they received feedback on whether their answer was correct or not after each answer, only the accuracy is comparable to the intrinsic rounds.

APPENDIX E: Incentive Effects

We conducted additional analysis focusing on the relationship between any incentive effects and the subjective measures. The new dependent variable was created by calculating the difference between the total correct in the extrinsic and intrinsic incentive schemes, only for the task the participants did first. As Table E1 shows, none of the subjective measures are statistically significantly different from zero.

Table E1: Incentive Effects

	(1) Incentive Effects	(2) Incentive Effects	(3) Incentive Effects	(4) Incentive Effects	(5) Incentive Effects
Age in months	-0.013	-0.013	-0.013	-0.012	-0.012
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Boy	0.194	0.198*	0.204*	0.196	0.194
	(0.120)	(0.118)	(0.118)	(0.119)	(0.122)
Fluid intelligence	0.116***	0.111**	0.111**	0.111**	0.111**
	(0.044)	(0.049)	(0.044)	(0.044)	(0.047)
Mouse use	0.001	0.002	0.005	-0.003	-0.000
	(0.040)	(0.040)	(0.041)	(0.042)	(0.042)
Conscientiousness	0.035				0.033
	(0.057)				(0.064)
Need for cognition		0.014			0.005
		(0.051)			(0.055)
Internal LOC			0.003		0.000
			(0.047)		(0.050)
External LOC			0.024		0.018
			(0.051)		(0.055)
Delay of gratification				0.080	0.080
				(0.108)	(0.112)
Constant	1.487	1.488	1.494	1.371	1.366
	(1.030)	(1.022)	(1.021)	(1.023)	(1.043)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0.312***	0.311***	0.311***	0.312***	0.312***
	(0.055)	(0.056)	(0.056)	(0.055)	(0.055)
sd(ε_{ij})	0.929**	0.930**	0.930**	0.930**	0.928**
	(0.034)	(0.034)	(0.034)	(0.033)	(0.033)
Observations	420	420	420	420	420

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

APPENDIX F: Reaction Times and Accuracy in the AX and Simon tasks

Tables F1 – F4 break down the results by reaction time and accuracy in the extrinsic condition of the AX and Simon tasks.[1] We find that children with a higher conscientiousness and need for cognition improve both their average reaction times and their accuracy in the AX task. Furthermore, participants with a higher need for cognition and an internal LOC decrease their reaction times in the Simon task.

Table F1: AX Reaction Time

	(1) Extrinsic reaction time	(2) Extrinsic reaction time	(3) Extrinsic reaction time	(4) Extrinsic reaction time
Age in months	0.002	0.002	0.002	0.002
	(0.007)	(0.006)	(0.007)	(0.007)
Boy	-0.401***	-0.414***	-0.442***	-0.428***
	(0.080)	(0.073)	(0.077)	(0.076)
Fluid intelligence	-0.240***	-0.204***	-0.221***	-0.229***
	(0.038)	(0.043)	(0.041)	(0.040)
Conscientiousness	-0.169**			
	(0.077)			
Need for cognition		-0.143***		
		(0.049)		
Internal LOC			-0.034	
			(0.048)	
External LOC			-0.059	
			(0.048)	
Delay of gratification				-0.000
				(0.083)
Constant	0.435	0.529	0.462	0.480
	(0.897)	(0.838)	(0.891)	(0.886)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0.071*	0.061*	0.077***	0.066**
	(0.097)	(0.088)	(0.075)	(0.084)
sd(ε_{ij})	0.888	0.892	0.899	0.903
	(0.077)	(0.080)	(0.080)	(0.080)
Observations	419	419	419	419

Standard errors in parentheses
** p < 0.10, ** p < 0.05, *** p < 0.01*

[1] Our sample sizes have decreased by one participant because two participants (one in each task) played a game in both rounds of the extrinsic task.

Table F2: Simon Reaction Time

	(1) Extrinsic reaction time	(2) Extrinsic reaction time	(3) Extrinsic reaction time	(4) Extrinsic reaction time
Age in months	-0.016*	-0.016**	-0.015*	-0.016**
	(0.008)	(0.007)	(0.008)	(0.008)
Boy	-0.727***	-0.718***	-0.747***	-0.731***
	(0.084)	(0.083)	(0.087)	(0.085)
Fluid intelligence	-0.236***	-0.197***	-0.229***	-0.230***
	(0.043)	(0.046)	(0.042)	(0.042)
Conscientiousness	-0.066			
	(0.059)			
Need for cognition		-0.199***		
		(0.057)		
Internal LOC			-0.108*	
			(0.062)	
External LOC			-0.024	
			(0.047)	
Delay of gratification				-0.114
				(0.077)
Constant	2.465**	2.529***	2.381**	2.625**
	(1.080)	(0.947)	(1.072)	(1.036)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.887**	0.868**	0.882**	0.888**
	(0.052)	(0.048)	(0.051)	(0.051)
Observations	418	418	418	418

Standard errors in parentheses
** p < 0.10, ** p < 0.05, *** p < 0.01*

Table F3: AX Accuracy

	(1) Extrinsic accuracy	(2) Extrinsic accuracy	(3) Extrinsic accuracy	(4) Extrinsic accuracy
Age in months	0.007	0.007	0.007	0.007
	(0.012)	(0.011)	(0.011)	(0.011)
Boy	-0.072	-0.048	-0.037	-0.039
	(0.082)	(0.080)	(0.082)	(0.082)
Fluid intelligence	0.075^{***}	0.051^{**}	0.058^{***}	0.063^{***}
	(0.022)	(0.020)	(0.022)	(0.021)
Conscientiousness	0.186^{*}			
	(0.103)			
Need for cognition		0.061^{*}		
		(0.036)		
Internal LOC			-0.037	
			(0.057)	
External LOC			0.030	
			(0.058)	
Delay of gratification				-0.064
				(0.098)
Constant	-1.220	-1.319	-1.318	-1.201
	(1.560)	(1.492)	(1.510)	(1.550)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0.001	0.034	0.040	0.046
	(0.049)	(0.184)	(0.365)	(0.137)
sd(ε_{ij})	0.959	0.974	0.973	0.975
	(0.176)	(0.184)	(0.182)	(0.183)
Observations	419	419	419	419

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table F4: Simon Accuracy

	(1) Extrinsic accuracy	(2) Extrinsic accuracy	(3) Extrinsic accuracy	(4) Extrinsic accuracy
Age in months	0.001	0.001	0.001	0.002
	(0.007)	(0.007)	(0.006)	(0.007)
Boy	-0.075	-0.072	-0.077	-0.074
	(0.121)	(0.121)	(0.120)	(0.122)
Fluid intelligence	0.109**	0.104*	0.113**	0.105**
	(0.054)	(0.056)	(0.054)	(0.053)
Conscientiousness	0.037			
	(0.074)			
Need for cognition		0.014		
		(0.047)		
Internal LOC			0.045	
			(0.050)	
External LOC			-0.049	
			(0.060)	
Delay of gratification				0.064
				(0.119)
Constant	0.177	0.178	0.221	0.086
	(0.832)	(0.813)	(0.802)	(0.815)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.984	0.985	0.983	0.985
	(0.094)	(0.094)	(0.093)	(0.095)
Observations	418	418	418	418

Standard errors in parentheses
** p < 0.10, ** p < 0.05, *** p < 0.01*

APPENDIX G: Sensitivity Analyses

Gender and Personality Traits

For most personality characteristics, there were no significant differences in the way they affect the effort provisions of boys and girls. However, for the synthetic effort measure we found that locus of control and greater delay of gratification had a positive effect on boys, but not on girls, for the intrinsic and the extrinsic conditions respectively. When examining the task-specific performances separately, this association was driven by the Simon task.

Table G1: Intrinsic effort and gender and personality interactions

	(1) Intrinsic effort	(2) Intrinsic effort	(3) Intrinsic effort	(4) Intrinsic effort
Age in months	0.017^{**}	0.017^{*}	0.014	0.017^{**}
	(0.009)	(0.009)	(0.009)	(0.009)
Fluid intelligence	0.039	0.033	0.026	0.037
	(0.032)	(0.035)	(0.032)	(0.032)
Mouse use	0.057^{*}	0.057^{*}	0.054^{*}	0.058^{*}
	(0.031)	(0.033)	(0.029)	(0.032)
Boy	-0.083	-0.097	-0.080	-0.245
	(0.098)	(0.094)	(0.099)	(0.178)
Conscientiousness	0.008			
	(0.081)			
Boy * Conscientiousness	0.092			
	(0.106)			
Need for cognition		-0.024		
		(0.066)		
Boy * Need for cognition		0.203^{*}		
		(0.109)		
Internal LOC			-0.053	
			(0.064)	
Boy * Internal LOC			0.163^{*}	
			(0.090)	
External LOC			0.114	
			(0.073)	
Boy * External LOC			-0.165	
			(0.110)	
Delay of gratification				-0.133
				(0.114)
Boy * Delay of gratification				0.231
				(0.225)
Constant	-2.193^{**}	-2.111^{**}	-1.764^{*}	-2.114^{**}
	(0.965)	(0.984)	(1.041)	(0.953)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0.314^{***}	0.313^{***}	0.314^{***}	0.313^{***}
	(0.052)	(0.052)	(0.054)	(0.052)
sd(ε_{ij})	0.923^{***}	0.918^{***}	0.918^{***}	0.925^{***}
	(0.028)	(0.027)	(0.028)	(0.027)
Observations	420	420	420	420

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table G2: Extrinsic effort and gender and personality interactions

	(1) Extrinsic effort	(2) Extrinsic effort	(3) Extrinsic effort	(4) Extrinsic effort
Age in months	0.009	0.009	0.009	0.011
	(0.009)	(0.008)	(0.008)	(0.008)
Fluid intelligence	0.251***	0.229***	0.241***	0.243***
	(0.044)	(0.042)	(0.040)	(0.039)
Mouse use	0.081**	0.084***	0.088**	0.077**
	(0.032)	(0.032)	(0.036)	(0.036)
Boy	0.323***	0.343***	0.363***	0.013
	(0.097)	(0.094)	(0.092)	(0.199)
Conscientiousness	0.167			
	(0.105)			
Boy * Conscientiousness	-0.049			
	(0.126)			
Need for cognition		0.048		
		(0.053)		
Boy * Need for cognition		0.154		
		(0.105)		
Internal LOC			0.036	
			(0.060)	
Boy * Internal LOC			-0.005	
			(0.102)	
External LOC			-0.007	
			(0.078)	
Boy * External LOC			0.085	
			(0.105)	
Delay of gratification				-0.121
				(0.149)
Boy * Delay of gratification				0.506**
				(0.232)
Constant	-1.709	-1.795*	-1.843*	-1.972**
	(1.076)	(1.018)	(1.048)	(0.986)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.886**	0.887**	0.894**	0.889**
	(0.048)	(0.049)	(0.050)	(0.047)
Observations	419	419	419	419

Standard errors in parentheses
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Sensitivity to Outlier Exclusion

We performed various robustness checks to analyze the sensitivity of results. First, we analyzed results excluding outliers. Outliers were defined as those individuals with accuracy below 50% for each AX and Simon task round, implying an exclusion of between 2% to 6.5% of the data. Results remain qualitatively similar for both intrinsic and extrinsic dependent variables. Second, we define outliers as those observations that are below the mean by 2.5 standard deviations for the AX and Simon tasks, and by 1.5 standard deviations for the Slider task.[2] Sample size is significantly reduced in this case, but results remain stable.

[2] There were no observations that were below the 2.5 standard deviation cut-off in the Slider task, which is why we took a stricter approach with a 1.5 standard deviation cut-off for that task.

Table G3: AX intrinsic excluding participants with accuracy below 50%

	(1) AX intrinsic	(2) AX intrinsic	(3) AX intrinsic	(4) AX intrinsic	(5) AX intrinsic
Age in months	0.018	0.018	0.019	0.020	0.020
	(0.015)	(0.016)	(0.015)	(0.016)	(0.017)
Boy	0.057	0.061	0.064	0.065	0.048
	(0.161)	(0.156)	(0.160)	(0.154)	(0.161)
Fluid intelligence	0.077	0.066	0.080	0.077*	0.082
	(0.049)	(0.054)	(0.055)	(0.047)	(0.057)
Conscientiousness	0.065				0.074
	(0.103)				(0.101)
Need for cognition		0.034			0.018
		(0.074)			(0.074)
Internal LOC			0.003		-0.024
			(0.062)		(0.060)
External LOC			-0.038		-0.039
			(0.077)		(0.083)
Delay of gratification				-0.139	-0.137
				(0.233)	(0.257)
Constant	-2.424	-2.404	-2.518	-2.551	-2.528
	(1.772)	(1.790)	(1.757)	(1.839)	(1.878)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0.289**	0.288**	0.294**	0.287**	0.290**
	(0.144)	(0.144)	(0.146)	(0.143)	(0.144)
sd(ε_{ij})	0.892***	0.894***	0.892***	0.893***	0.887***
	(0.038)	(0.036)	(0.036)	(0.031)	(0.035)
Observations	181	181	181	181	181

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table G4: AX extrinsic excluding participants with accuracy below 50%

	(1) AX extrinsic	(2) AX extrinsic	(3) AX extrinsic	(4) AX extrinsic	(5) AX extrinsic
Age in months	0.000	0.000	-0.000	0.001	0.001
	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Boy	0.117	0.143	0.156	0.147	0.104
	(0.113)	(0.113)	(0.111)	(0.115)	(0.114)
Fluid intelligence	0.185***	0.153***	0.168***	0.168***	0.175***
	(0.037)	(0.035)	(0.037)	(0.034)	(0.041)
Conscientiousness	0.219**				0.219**
	(0.094)				(0.102)
Need for cognition		0.093***			0.058
		(0.032)			(0.042)
Internal LOC			0.020		-0.022
			(0.056)		(0.063)
External LOC			0.015		-0.012
			(0.055)		(0.061)
Delay of gratification				0.142	0.148
				(0.107)	(0.112)
Constant	-0.708	-0.800	-0.764	-0.992	-0.972
	(0.802)	(0.654)	(0.680)	(0.609)	(0.728)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.895	0.914	0.918	0.916	0.889
	(0.110)	(0.112)	(0.112)	(0.112)	(0.109)
Observations	410	410	410	410	410

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table G5: Simon intrinsic excluding participants with accuracy below 50%

	(1) Simon intrinsic	(2) Simon intrinsic	(3) Simon intrinsic	(4) Simon intrinsic	(5) Simon intrinsic
Age in months	0.008	0.007	0.009	0.009	0.008
	(0.011)	(0.012)	(0.011)	(0.013)	(0.015)
Boy	0.032	0.044	0.009	0.012	0.062
	(0.291)	(0.264)	(0.332)	(0.269)	(0.392)
Fluid intelligence	-0.072	-0.080	-0.052	-0.062	-0.092
	(0.080)	(0.134)	(0.129)	(0.103)	(0.146)
Conscientiousness	-0.075				-0.127
	(0.140)				(0.166)
Need for cognition		0.139			0.137
		(0.128)			(0.126)
Internal LOC			0.094		0.096
			(0.144)		(0.149)
External LOC			-0.018		-0.010
			(0.202)		(0.231)
Delay of gratification				0.075	0.022
				(0.145)	(0.179)
Constant	-0.822	-0.770	-0.906	-1.069	-0.875
	(1.514)	(1.613)	(1.543)	(1.883)	(2.065)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.933	0.928	0.931	0.937	0.916
	(0.094)	(0.079)	(0.077)	(0.091)	(0.074)
Observations	66	66	66	66	66

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table G6: Simon extrinsic excluding participants with accuracy below 50%

	(1) Simon extrinsic	(2) Simon extrinsic	(3) Simon extrinsic	(4) Simon extrinsic	(5) Simon extrinsic
Age in months	0.015	0.015	0.015	0.015	0.015
	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
Boy	0.311***	0.301***	0.321***	0.317***	0.290***
	(0.086)	(0.086)	(0.089)	(0.089)	(0.086)
Fluid intelligence	0.166***	0.124***	0.168***	0.162***	0.132***
	(0.044)	(0.048)	(0.043)	(0.043)	(0.047)
Conscientiousness	0.064				0.018
	(0.080)				(0.083)
Need for cognition		0.202***			0.198***
		(0.040)			(0.048)
Internal LOC			0.066		0.023
			(0.052)		(0.052)
External LOC			-0.030		-0.051
			(0.062)		(0.063)
Delay of gratification				0.048	0.070
				(0.054)	(0.062)
Constant	-2.387	-2.338	-2.320	-2.439*	-2.453*
	(1.505)	(1.438)	(1.473)	(1.480)	(1.372)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.893	0.874	0.892	0.895	0.869*
	(0.074)	(0.072)	(0.074)	(0.074)	(0.071)
Observations	393	393	393	393	393

Standard errors in parentheses
** p < 0.10, ** p < 0.05, *** p < 0.01*

Table G7: Intrinsic effort excluding participants below total correct cut-off

	(1) Intrinsic effort	(2) Intrinsic effort	(3) Intrinsic effort	(4) Intrinsic effort	(5) Intrinsic effort
Age in months	0.017*	0.018*	0.017*	0.017*	0.017*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Boy	-0.091	-0.092	-0.080	-0.083	-0.091
	(0.105)	(0.103)	(0.105)	(0.106)	(0.108)
Fluid intelligence	0.035	0.021	0.031	0.034	0.023
	(0.030)	(0.031)	(0.031)	(0.030)	(0.032)
Mouse use	0.049	0.050	0.052*	0.050	0.053
	(0.031)	(0.031)	(0.030)	(0.032)	(0.034)
Conscientiousness	0.039				0.022
	(0.064)				(0.065)
Need for cognition		0.070			0.063
		(0.043)			(0.040)
Internal LOC			0.025		0.004
			(0.047)		(0.047)
External LOC			0.017		0.011
			(0.053)		(0.056)
Delay of gratification				-0.028	-0.028
				(0.108)	(0.110)
Constant	-2.169**	-2.208**	-2.185**	-2.152**	-2.159**
	(1.034)	(1.049)	(1.038)	(1.017)	(1.053)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0.330***	0.331***	0.331***	0.332***	0.330***
	(0.055)	(0.054)	(0.055)	(0.055)	(0.054)
sd(ε_{ij})	0.919***	0.917***	0.919***	0.920***	0.916***
	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)
Observations	408	408	408	408	408

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table G8: Extrinsic effort excluding participants below total correct cut-off

	(1) Extrinsic effort	(2) Extrinsic effort	(3) Extrinsic effort	(4) Extrinsic effort	(5) Extrinsic effort
Age in months	0.012	0.010	0.011	0.012	0.012
	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
Boy	0.352^{***}	0.354^{***}	0.384^{***}	0.368^{***}	0.329^{***}
	(0.090)	(0.093)	(0.087)	(0.093)	(0.098)
Fluid intelligence	0.264^{***}	0.222^{***}	0.254^{***}	0.256^{***}	0.225^{***}
	(0.058)	(0.057)	(0.059)	(0.055)	(0.059)
Mouse use	0.091^{**}	0.091^{**}	0.099^{**}	0.084^{**}	0.078^{**}
	(0.041)	(0.037)	(0.041)	(0.040)	(0.040)
Conscientiousness	0.147^{*}				0.121
	(0.087)				(0.097)
Need for cognition		0.165^{***}			0.146^{**}
		(0.062)			(0.064)
Internal LOC			0.044		-0.001
			(0.057)		(0.059)
External LOC			0.035		0.007
			(0.075)		(0.075)
Delay of gratification				0.153^{*}	0.172^{*}
				(0.089)	(0.092)
Constant	-2.226^{*}	-2.144^{*}	-2.187^{*}	-2.415^{**}	-2.401^{**}
	(1.278)	(1.245)	(1.226)	(1.168)	(1.212)
Controlled for task order	YES	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES	YES
sd(u_j)	0	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.893^{*}	0.890^{*}	0.901^{*}	0.902^{*}	0.878^{**}
	(0.054)	(0.056)	(0.055)	(0.054)	(0.053)
Observations	332	332	332	332	332

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Sensitivity to Including Teacher Observations

For a further sensitivity check we include controls for teacher observations about difficulties observed in children in the regressions together with all personality variables. Results do not change qualitatively following this inclusion.

Table G9: Intrinsic effort controlling for teacher evaluation

	(1) Intrinsic effort	(2) Slider intrinsic	(3) AX intrinsic	(4) Simon intrinsic
Age in months	0.017^{**}	0.025	0.015	0.007
	(0.009)	(0.018)	(0.013)	(0.014)
Boy	-0.095	-0.285^{**}	0.045	0.079
	(0.098)	(0.114)	(0.143)	(0.366)
Fluid intelligence	0.023	0.030	0.092	-0.100
	(0.034)	(0.034)	(0.061)	(0.126)
Mouse use	0.059^{*}	0.064		
	(0.034)	(0.059)		
Conscientiousness	0.041	0.031	0.113	-0.134
	(0.064)	(0.104)	(0.100)	(0.176)
Need for cognition	0.054	0.055	0.020	0.155
	(0.041)	(0.074)	(0.074)	(0.102)
Internal LOC	-0.003	0.033	-0.052	0.073
	(0.047)	(0.089)	(0.050)	(0.173)
External LOC	0.025	0.088	0.003	-0.003
	(0.056)	(0.070)	(0.090)	(0.231)
Delay of gratification	-0.030	0.026	-0.104	-0.015
	(0.108)	(0.095)	(0.251)	(0.176)
Teacher observation: ADHD	0.040	-0.110	0.352	0.027
	(0.150)	(0.270)	(0.309)	(0.265)
Teacher observation: other	-0.159	-0.194	-0.354	0.075
	(0.130)	(0.240)	(0.298)	(0.175)
Constant	-2.116^{**}	-2.863	-1.996	-0.737
	(0.975)	(2.282)	(1.510)	(1.934)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0.315^{***}	0.274^{***}	0.239^{***}	0
	(0.052)	(0.105)	(0.133)	(0.000)
sd(ε_{ij})	0.921^{***}	0.910^{**}	0.901^{**}	0.910
	(0.028)	(0.039)	(0.041)	(0.079)
Observations	420	164	186	70

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Table G10: Extrinsic effort controlling for teacher evaluation

	(1) Extrinsic effort	(2) Slider extrinsic	(3) AX extrinsic	(4) Simon extrinsic
Age in months	0.012	0.013	0.002	0.011
	(0.008)	(0.008)	(0.007)	(0.008)
Boy	0.367***	0.430***	0.144	0.228**
	(0.101)	(0.069)	(0.097)	(0.114)
Fluid intelligence	0.209***	0.171***	0.143***	0.159***
	(0.048)	(0.052)	(0.034)	(0.046)
Mouse use	0.064*	0.095**		
	(0.034)	(0.037)		
Conscientiousness	0.118	-0.013	0.252**	0.016
	(0.093)	(0.102)	(0.116)	(0.091)
Need for cognition	0.077	0.017	0.041	0.114**
	(0.051)	(0.055)	(0.041)	(0.051)
Internal LOC	-0.004	0.005	-0.058	0.052
	(0.053)	(0.053)	(0.062)	(0.047)
External LOC	0.017	0.074	0.008	-0.057
	(0.063)	(0.056)	(0.060)	(0.063)
Delay of gratification	0.114	0.106	0.044	0.108
	(0.084)	(0.076)	(0.108)	(0.094)
Teacher observation: ADHD	-0.495*	-0.382***	-0.553*	-0.163
	(0.255)	(0.140)	(0.320)	(0.266)
Teacher observation: other	-0.498**	-0.616***	-0.290	-0.209
	(0.238)	(0.231)	(0.202)	(0.180)
Constant	-2.144**	-2.247**	-0.790	-1.648
	(0.968)	(1.036)	(0.979)	(1.062)
Controlled for task order	YES	YES	YES	YES
Controlled for neighb. income	YES	YES	YES	YES
sd(u_j)	0	0	0	0
	(0.000)	(0.000)	(0.000)	(0.000)
sd(ε_{ij})	0.866***	0.903***	0.886	0.929
	(0.045)	(0.033)	(0.095)	(0.057)
Observations	419	420	420	419

Standard errors in parentheses
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$*

Analyses using Data without Imputations

Finally, we perform the reported analyses from the article on the data without imputations and find stable results. We conclude that results presented here are robust to data imputations.

Table G11: Results of analyses by task for the intrinsic and extrinsic conditions

	(1) Effort	(2) Slider	(3) AX	(4) Simon
Intrinsic incentive scheme				
Conscientiousness	0.020	0.040	0.030	-0.005
	(0.046)	(0.100)	(0.062)	(0.033)
Observations	415	163	182	70
Need for cognition	0.071*	0.090	0.043	0.141
	(0.040)	(0.076)	(0.068)	(0.098)
Observations	415	162	183	70
Internal LOC	0.003	-0.022	-0.016	0.089
	(0.055)	(0.116)	(0.030)	(0.181)
External LOC	-0.007	-0.002	-0.005	-0.035
	(0.055)	(0.075)	(0.074)	(0.237)
Observations	281	108	124	49
Delay of gratification	-0.030	0.032	-0.095	0.061
	(0.107)	(0.094)	(0.231)	(0.128)
Observations	416	163	183	70
Extrinsic incentive scheme				
Conscientiousness	0.118***	0.043	0.073**	0.142***
	(0.036)	(0.048)	(0.033)	(0.036)
Observations	414	415	415	414
Need for cognition	0.113**	0.042	0.079**	0.129**
	(0.052)	(0.058)	(0.032)	(0.051)
Observations	414	415	415	414
Internal LOC	0.015	0.009	-0.042	0.064
	(0.058)	(0.060)	(0.065)	(0.057)
External LOC	0.031	0.074	0.050	-0.061
	(0.069)	(0.059)	(0.056)	(0.076)
Observations	281	281	281	281
Delay of gratification	0.114	0.117*	0.059	0.084
	(0.080)	(0.067)	(0.101)	(0.087)
Observations	415	416	416	415

Note: Models estimated with each scale separately (except internal and external locus of control together), controlling for: sex, age, fluid intelligence, neighborhood income, task order, and mouse use (for models (1) and (2) only).
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$